Journal of
Computer & Electrical and Electronics Engineering Sciences
Original Article
DOI: 10.51271/JCEEES-0009

Classification of heart disease dataset with k-NN optimized by pso and gwo algorithms

Murat Lüy[1] Nuri Alper Metin[2]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Kırıkkale Univercity, Kırıkkale, Turkey
[2]Department of Electronic Communication Program, Kırıkkale Vocational School, Kırıkkale Univercity, Kırıkkale, Turkey

Cite this article: Lüy M, Metin NA. Classification of heart disease dataset with k-NN optimized by pso and gwo algorithms. *J Comp Electr Electron Eng Sci.* 2023;1(2):41-45

Corresponding Author: Nuri Alper Metin, nurialpermetin@kku.edu.tr

Received: 12/08/2023 ◆ **Accepted**: 06/10/2023 ◆ **Published**: 31/10/2023

ABSTRACT

Lifestyle changes worldwide are increasing chronic diseases (CD). This research concluded that cardiovascular diseases (CVDs) cause 46% of global mortality, excluding communicable diseases and accidents, and heart attacks are 7.4 million of the 17.5 million CVD deaths. 2030 will see 22.2 million cardiovascular deaths. Avoiding and treating most HD reduces cardiovascular disease fatalities. In this study, the data set for heart disease is optimized with PSO and GWO, and classification is performed with k-NN.

Keywords: PSO, GWO, k-NN

INTRODUCTION

Today, viral cardiovascular disease creates heart issues quickly. If the heart fails, all bodily functions will be impacted. One-third of the world dies of heart failure, according to WHO figures. 2016 marks 18 million CVD sufferers. Heart failure or attack killed almost 84%. Down nations had three-quarters of cardiovascular disease mortality. 2015 survey. He has non-infectious heart failure from 82 to 20. From these statistics, CVD caused her heart illness (Dulhare, 2018; Hasanova et al., 2022; Naga & Asst, 2023; Sengur, 2008; Seslier & Karakuş, 2022; Tharwat et al., 2018). Health risks, including smoking, poor diet, and excessive alcohol usage, often cause cardiovascular disease deaths. Advanced cardiovascular disease or CVD patients, diabetes, hypertension, hyperlipidemia, and other risk factors should be discussed early on. Adipose-stored CVD puffs have formed inside atherosclerosis and blood clusters. It harms the eyes, kidneys, heart, and mind. UK/US CVD has killed severely—cardiac stroke. Episodes are severe events that stop heart or blood circulation. CVD development targets fat buildup and venous consequences. Obesity, cigarette usage, and poor nutrition contribute to CVD strokes and failures. The heart circulates blood via blood vessels. Blood flow eliminates metabolic waste from the heart and provides nutrients and oxygen to various areas. Poor blood flow may damage organs and worsen heart failure. This study presents PSO and GWO optimization and k-NN classification for heart disease (Dulhare, 2018; Hasanova et al., 2022; Naga & Asst, 2023; Sengur, 2008; Seslier & Karakuş, 2022; Tharwat et al., 2018).

Literature Review

Khourdifi et al., Machine learning predict heart disease. Optimization strategies may handle complex non-linear problems. FCBF filtered redundant characteristics to classify heart disease in this article. k-NN, Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and a PSO-ACO-optimized Multilayer Perception, Artificial Neural Network Classification. A heart disease dataset tests the hybrid heart disease classification method. Its efficacy and robustness were shown. This study examines machine learning approaches using accuracy, precision, recall, f1-score, etc. Improved FCBF, PSO, and ACO models classify 99.65% accurately. The suggested technique surpasses classification (Khourdifi & Bahaj, 2019).

Sandhiya et al., Heart disease kills all ages since people don't know their severity. Heart disease kinds and monitoring are crucial in our fast-paced world. IoT and Deep Learning create a patient-safe heart disease monitoring system. Feature selection helps deep learning classification. Our proposed system monitors heart disease using IoT device inputs. It classifies individuals by heart illness kind and severity. Finally, heart disease kind and inputs notify patients. The recommended system predicted better testing (UPalani Teaching Fellow Professor, 2022).

Tama et al., Coronary heart disease (CHD) are common and severe. Poor lives kill the most people globally. Since cardiac attacks are symptomless, sophisticated detection is needed. This article introduces classifier ensembles for CHD detection. Two-tier ensemble classifiers are ensemble-specific. A stacked design uses random forest class label prediction, gradient boosting, and extreme slope boosting. Z-Alizadeh Sani, Statlog, Cleveland, Ohio, and Hungarian heart disease datasets validate the detection model. Particle swarm optimization chooses each dataset's most significant

This work is licensed under a Creative Commons Attribution 4.0 International License.

properties. Finally, a two-fold statistical study disproves classifier performance disparities based on assumptions. Our 10-fold cross-validation outperforms ensemble base classifiers. Our detection system surpassed classifier groups and classifiers in accuracy and AUC. This study reveals that our model contributes more than earlier publications (Tama et al., 2020).

Khourdifi et al., The hybrid Machine Learning model for the proposed PA-RF, a classification based on the Random Forest model, was optimized by PSO, ACO, and FCBF to filter redundant and irrelevant characteristics to improve heart disease classification. Heart disease data is mixed. The hybrid method is effective and durable in classifying heart disease using different data sources. Thus, this study evaluates autonomous learning algorithms using Accuracy, Precision, Recall, F1-Score, etc. This study used UCI's autonomous learning repository's "Heart Disease" data set. PA-RF excels (Khourdifi & Bahaj, 2019).

The study conducted by Muthukaruppan et al. focuses on developing a Particle Swarm Optimization (PSO)-based fuzzy expert system for Computer-Aided Design (CAD) diagnosis. The system was influenced by the Cleveland and Hungarian Heart Disease databases. Using a decision tree (DT) enabled the identification of diagnostic characteristics within datasets encompassing many input qualities. The DT output was enhanced by transforming it into crisp if-then rules and incorporating a fuzzy rule base, which was further optimized using PSO to tune the Fuzzy Membership Functions (FMFs). The fuzzy expert system achieved a classification accuracy of 93.27% using improved membership functions. This methodology demonstrates enhanced interpretability of fuzzy expert system selections (Muthukaruppan & Er, 2012).

Syafi et al., The heart pumps blood. Data mining from a massive data warehouse aids decision-making. Data structure precedes data mining. Then, Information Gain Ratio and Particle Swarm Optimization choose the best features. Adaboost maximized accuracy. Data classification follows. C4.5 classes. The C4.5 approach uses Information Gain Ratio (IGR) and Particle Swarm Optimization, then applying the Adaboost ensemble has a 96.68% accuracy rate, whereas the one without feature selection has 95.87%. Information Gain Ratio and Particle Swarm Optimization utilizing the Adaboost ensemble may improve C4.5 classification algorithm performance (Qois Syafi, 2022).

Roostaee et al., Large heart disease data sets challenge analysis. All features disappoint. Thus, qualities need improvement. Binary cuckoo optimization reduces property. SVM classifies heart disease best. This article simplifies illness diagnosis. Accuracy, sensitivity, and specificity. 14 attributes for 303 ML Repository users. Current approaches are costly, accurate, and time-consuming. The suggested technique is faster, more accurate, and more efficient (Roostaee & Ghaffary, 2016).

Li et al., Electrocardiograms (ECGs) objectively measure heart function and physiological condition, making them helpful in diagnosing cardiac illness. Feature extraction determines AECG judgment accuracy. Environmental noise makes ECG detection challenging. Different AECG species exist. ECG findings from a long time ago cannot be utilized to identify or diagnose sickness. Thus, AECG detection requires an innovative classification method with exact feature extraction. This study used ECG feature extraction to create an AECG detection and cardiac disease diagnostic algorithm (Li et al., 2021).

Dubey et al., Cardiovascular diseases are global health issues. Early diagnosis may improve survival for cardiovascular illnesses, which have the highest worldwide mortality rate and increase with age. ML and optimization predict heart diseases in this paper. ML approaches were used to classify data and IACPSO to choose the best features. UCI ML evaluated Cleveland, Statlog, and Hungarian heart disease datasets. Model performance was assessed. LR and SVMGS outperformed on Statlog, Cleveland, and Hungarian datasets. IACPSO-ML improved performance indicators by 3–33% (Dubey et al., 2022).

Jabbar et al., Knowledgeable information has been mined extensively using data mining methods from medical databases. Classification is a kind of supervised learning used in data mining, and it is put to use by building models that characterize various classes of data. The nearest neighbor (k-NN) method is the simplest, most widely used, and most successful pattern recognition algorithm. Classifying samples according to the category of their closest neighbor, k-NN is a simple classifier. The amount of data in healthcare databases is quite large. Classification performance may suffer if the data set includes superfluous or unnecessary characteristics. In INDIA, heart disease is number one among all causes of mortality. Heart disease is the top cause of death in Andhra Pradesh, accounting for 32% of all fatalities. It aligns with the 35% seen in Canada and the USA. Creating a decision support system is crucial to help physicians decide whether to take preventive steps. This research combines KNN with a genetic algorithm to provide a fresh approach to accurate categorization. A global search conducted by genetic algorithms over vast, multimodal landscapes may find the optimal solution. The experimental data indicate that our method enhances the diagnostic accuracy of heart illness (Jabbar et al., 2013).

This study involved the utilization of PSO and GWO algorithms for the optimization of the data set. Additionally, the classification task was carried out using the k-NN algorithm. The accuracy and processing time of both algorithms are compared.

Table 1. Literature review

Literature Articles	Optimization Methods	Classification Methods
Khourdifi et al (2018)	ANN-PSO, ANN-ACO	k-NN, SVM, RF, NB, MLP
Sandhiya et al. (2022)	GWO	DBN
Tama et al. (2020)	Two-Tier PSO	RF, GBM, XGBoost
Khourdifi et al (2019)	RF-PSO, RF ACO	k-NN, SVM, RF, NB, MLP, PA-RF
Muthukaruppan et al. (2012)	-	PSO-Fuzzy
Syafi et al. (2022)	PSO, IGR	C4.5
Roostaee et al. (2016)	BCOA	SVM
Li et al. (2021)	PSO-BPNN	PCA
Dubey et al. (2022)	IAPSO	LR, DT, RF, SVM, SVMGS k-NN, NB
Jabbar et al. (2013)	GA	k-NN

This research aims to enhance classification accuracy using an optimized k-NN algorithm to diagnose heart disease. PSO and GWO algorithms are used to optimize the heart disease dataset.

METHODS

In this study, PSO and GWO are used for data set optimization. The PSO algorithm is an example of a heuristic algorithm that works by repeatedly attempting to improve upon an existing solution to a problem. The animals' interactions with one another provided the basis for this program. A parameter value is specified in the PSO method to maximize the size of the particles scattered across the search space and the particles wandering the search space. The direction of these particles in the search space is determined not only by their flight route but also by the collective flight path of the flock, just as it is in a flock of birds (Wadhawan & Maini, 2022). The Block Diagram of PSO is shown in Figure 1.

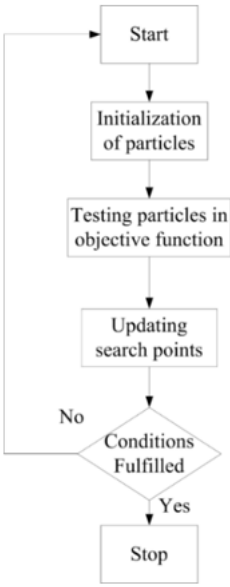

Figure 1. Block diagram of PSO

The grey wolf optimization algorithm, often known as the GWO, is a technique that mathematically models and imitates the behavior of a pack of grey wolves' behavior when hunting. In the hierarchy of wolf packs, grey wolves are referred to as the alpha (∝), beta (β), delta (δ), and omega (ω) positions, respectively. Because it plays by the same rules as the other species, the alpha wolf pack is considered the head of the pack. The deputy leader is referred to as the beta wolf, and they assist the alpha wolf throughout the decision-making process. The grey wolves designated as omega are positioned at the bottom of this hierarchy. The wolf is known as a delta if it does not belong to any species stated in the pack. An intriguing illustration of how grey wolves communicate with one another is shown in the behavior of group hunting. The behavior of GWOs may be broken down into four distinct stages: searching, surrounding, assaulting, and hunting (Al-Tashi et al., 2019; Mirjalili et al., 2014). The Block Diagram of GWO is shown in Figure 2.

K-NN classifier is a well-known classification method that is also relatively straightforward. Fix and Hodges initially presented it as a non-parametric approach, meaning it does not make any assumptions about the input data distribution. As a result, it has found widespread usage in various applications since its beginning.

The K-Nearest Neighbors (k-NN) classifier assigns an unidentified sample to a specific category by evaluating the distance between the sample and all labeled instances.

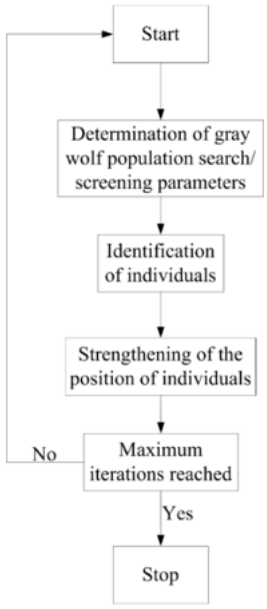

Figure 2. Block diagram of GWO

The unidentifiable sample is allocated to the class with the most samples within its k-nearest neighbors. The k-NN classifier algorithm relies on three essential components. Choosing a precise value for the integer k is crucial to deciding how many closest neighbors should be considered throughout the analysis. Additionally, it is imperative to have training data that includes labeled samples, as these can be adjusted by adding or eliminating samples. Finally, a distance metric is utilized to compute the closeness or proximity between different data points. The utilization of the Euclidean distance metrics to calculate the distance between samples is exemplified in Equation (1) within the k-NN algorithm. The k-NN classifier is advantageous due to its traceability and ease of construction. Nevertheless, storing every sample used for training in memory during runtime requires categorizing it as a memory-based classification approach (Karabulut et al., 2019; Tharwat et al., 2018; Prasath et al., 2017).

$$d(x_i, x_j) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (1)$$

Datasets Description

The dataset under consideration originates from 1988 and comprises four distinct databases: Cleveland, Hungary, Switzerland, and Long Beach V. The dataset comprises 76 attributes, encompassing the predicted attribute. However, all experiments published thus far exclusively utilize a subset of 14 attributes. The term "target" pertains to the existence of cardiovascular disease in the individual. The variable is assigned integer values, where 0 represents the absence of disease, and 1 represents the presence of disease.

The heart disease dataset is defined by gender, age, pain in the chest type (4 values), laying down level of blood pressure, cholesterol levels in mg/dl fasting blood, sugar>120 mg/dl, maximum heart rate accomplished, exercise-induced chest discomfort, old maximum=ST depression caused by physical activity compared to a resting state, the slope of the highest point during exercise, the segment of the ST with three or more significant vessels (0-3) colored by fluoroscopy that,

and other variables are defined in the database. Typical is 0, fixed defects are 1, and reversible defects are 2 (Kaggle, A.D: 02/08/2023).

Predictive Methods

Early detection of the disease and appropriate treatment are essential in preventing the increasing number of male and female patients with heart disease and reducing the risk of death. Today, the use of optimization and classification methods for the causes of heart disease is increasing. With the optimization and classification methods used, it is ensured that accurate results are obtained, the time required for disease detection is minimized, and human errors are prevented. This study used particle swarm optimization (PSO) and grey wolf optimization (GWO) algorithms to optimize the data set. The k-Nearest Neighbors (k-NN) algorithm also classified the data set. The accuracy and processing time of both algorithms are compared.

RESULTS

PSO parameters are given in Table 2. PSO consists of parameters lb (lower bound), ub (upper bound), coefficients c1, c2, and w (weight).

Table 2. PSO parameters

Parameter Name	Parameter Value
lb	0
ub	1
c1	2
c2	2
w	0.9
Number of solutions	10
Iteration number	1000

GWO parameters are given in Table 3. GWO consists of parameters lb (lower bound), ub (upper bound), coefficients c1, c2, and c3.

Table 3. GWO parameters

Parameter Name	Parameter Value
lb	0
ub	1
c1	Between 0 and 2
c2	Between 0 and 2
c3	Between 0 and 2
Number of solutions	10
Iteration number	1000

This paper optimizes the heart disease dataset with PSO and GWO. Fitness values according to iteration are shown in Figure 3.

The processing time of PSO and GWO optimization algorithms is given in Table 4.

Table 4. The processing time of PSO and GWO

Optimization Type	Processing Time (sec)
GWO	60.120612
PSO	60.768583

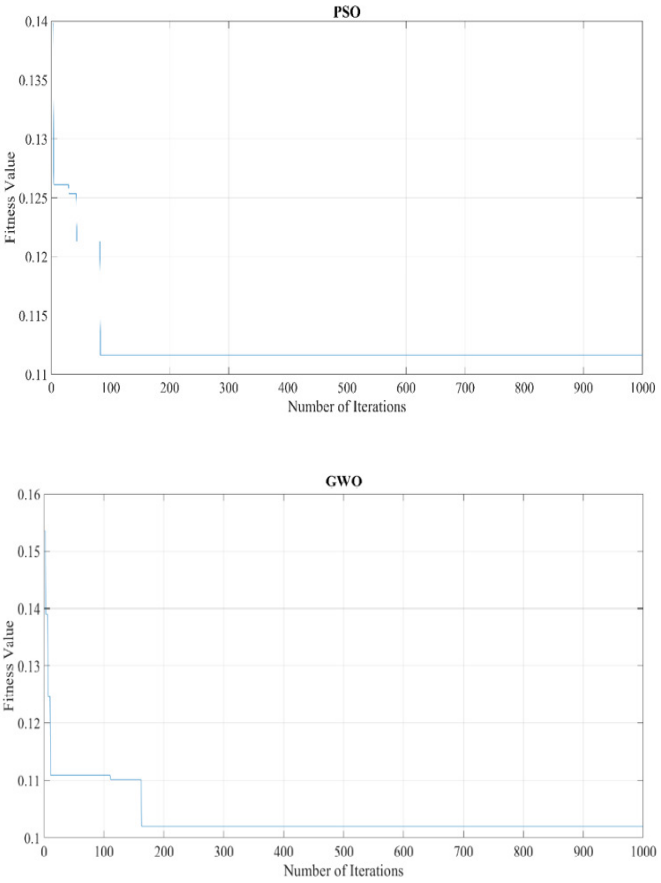

Figure 3. Fitness values according to the number of iterations of GWO and PSO

For k-NN classification on the GWO and PSO optimized heart disease dataset, k is chosen to be 5. accuracy is shown in Figure 4.

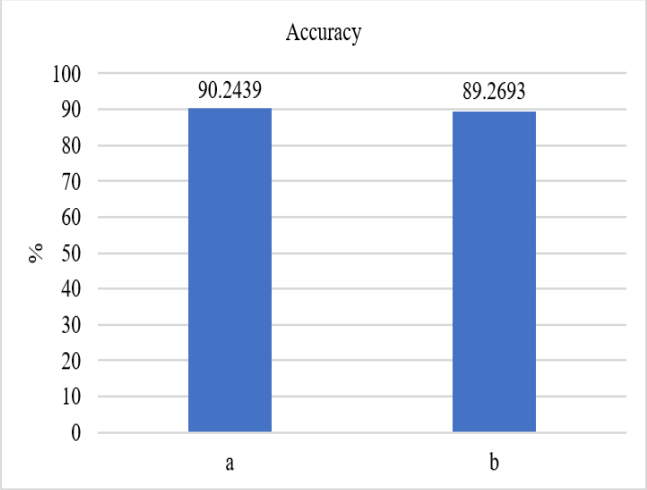

Figure 4. a) Accuracy calculated using GWO in heart disease dataset optimization b) Accuracy calculated using PSO in heart disease dataset optimization

DISCUSSION

This study optimized the heart disease dataset with GWO and PSO, and classification was performed with k-NN. The comparison process was performed in a Matlab environment. GWO proved to be better than PSO in both processing time and accuracy. The proposed system was compared with studies in the literature in terms of accuracy. This situation is given in Table 5.

Table 5. A comparison of accuracy

Study	Feature Selection	Classify Techniques	Disease	Accuracy(%)
Khourdifi et al. (2018)	ANN-PSO, ANN-ACO	k-NN	Heart Disease	99.65
Sandhiya et al. (2022)	NGWO	DBN	Heart Disease	85.4
Tama et al. (2020)	Two-Tier PSO	RF, GBM, XGBoost	Heart Disease	85.71
Khourdifi et al. (2019)	PSO, ACO	k-NN	Heart Disease	84.13
Muthukaruppan et al. (2012)	-	PSO-Fuzzy	Heart Disease	93.27
Syafi et al. (2022)	PSO-IGR	C 4.5	Heart Disease	96.68
Roostaee et al. (2016)	BCOA	SVM	Heart Disease	84.44
Li et al. (2021)	PSO-BPNN	PCA	ECG	96
Dubey et al. (2022)	IACPSO	k-NN	Heart Disease	92
Jabbar	GA	k-NN	Heart Disease	95.73
Proposed System	PSO, GWO	k-NN	Heart Disease	90.2439 (GWO) 89.2693 (PSO)

For future studies, the system's accuracy can be increased by changing the optimization algorithm and classifier in the system.

CONCLUSION

This study uses PSO and GWO algorithms for heart disease dataset optimization. The k-NN algorithm classifies the optimized datasets. These algorithms are run for 1000 iterations. When the heart disease dataset is optimized with the PSO algorithm and classified with k-NN, it has 60.768583 processing time and 89.2683% accuracy rate, while when it is optimized with GWO and classified with k-NN, it has 60.120612 processing time and 90.2439% accuracy rate. As a result of the simulation analysis shows that the GWO algorithm has a shorter processing time and higher accuracy rate than PSO when the heart disease dataset is optimized and classified with k-NN.

REFERENCES

1. Al-Tashi, Q., Rais, H., & Jadid, S. (2019). Feature selection method based on grey wolf optimization for coronary artery disease classification. Advances in Intelligent Systems and Computing, 843, 257-266. https://doi.org/10.1007/978-3-319-99007-1_25
2. Dubey, A. K., Sinhal, A. K., & Sharma, R. (2022). An Improved Auto Categorical PSO with ML for Heart Disease Prediction. *Engineering, Technology & Applied Science Research*, 12(3), 8567-8573. https://doi.org/10.48084/etasr.4854
3. Dulhare, U. N. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Biomedical Research*, 29(12), 2646-2649. https://doi.org/10.4066/biomedicalresearch.29-18-620
4. Hasanova, H., Tufail, M., Baek, U. J., Park, J. T., & Kim, M. S. (2022). A novel blockchain-enabled heart disease prediction mechanism using machine learning. *Computers and Electrical Engineering,* 101(May), 108086. https://doi.org/10.1016/j.compeleceng.2022.108086
5. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm. *Procedia Technology,* 10, 85-94. https://doi.org/10.1016/j.protcy.2013.12.340
6. Karabulut, B., Arslan, G., & Ünver, H. M. (2019). A Weighted Similarity Measure for k-Nearest Neighbors Algorithm. Celal Bayar Üniversitesi Fen Bilimleri Dergisi, 15(4), 393-400. https://doi.org/10.18466/cbayarfbe.618964
7. Khourdifi, Y., & Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *International Journal of Intelligent Engineering and Systems,* 12(1), 242-252. https://doi.org/10.22266/ijies2019.0228.24
8. Khourdifi, Y., & Bahaj, M. (2019). The Hybrid Machine Learning Model Based on Random Forest Optimized by PSO and ACO for Predicting Heart Disease. https://doi.org/10.4108/eai.24-4-2019.2284088
9. Li, G., Tan, Z., Xu, W., Xu, F., Wang, L., Chen, J., & Wu, K. (2021). A particle swarm optimization improved BP neural network intelligent model for electrocardiogram classification. BMC Medical Informatics and Decision Making, 21(Suppl 2), 1-15. https://doi.org/10.1186/s12911-021-01453-6
10. Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. Advances in Engineering Software, 69, 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007
11. Muthukaruppan, S., & Er, M. J. (2012). A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. *Expert Systems with Applications,* 39(14), 11657-11665. https://doi.org/10.1016/j.eswa.2012.04.036
12. Naga, M., & Asst, S. (2023). Detection of Cardiovascular Disease using Machine Learning, Genetic Algorithms and Particle Swarm Optimization. *IJERT,* 12(03), 120-127.
13. Prasath, V. B. S., Alfeilat, H. A. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., & Salman, H. S. E. (2017). Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier -- A Review. 1-39. https://doi.org/10.1089/big.2018.0175
14. Qois Syafi, M. (2022). Increasing Accuracy of Heart Disease Classification on C4.5 Algorithm based on information gain ratio and particle swarm optimization using adaboost ensemble. *Journal of Advances in Information Systems and Technology*, 4(1), 100-112. https://journal.unnes.ac.id/sju/index.php/jaist
15. Roostaee, S., & Ghaffary, H. R. (2016). Diagnosis of heart disease based on meta heuristic algorithms and clustering methods. *Journal of Electrical and Computer Engineering Innovations JECEI,* 4(2), 105-110. https://doi.org/10.22061/jecei.2016.570
16. Sengur, A. (2008). An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases. *Computers in Biology and Medicine,* 38(3), 329-338. https://doi.org/10.1016/j.compbiomed.2007.11.004
17. Seslier, T., & Karakuş, M. Ö. (2022). In healthcare applications of machine learning algorithms for prediction of heart ATTACKS. *Journal of Scientific Reports-A*, 051, 358-370.
18. Tama, B. A., Im, S., & Lee, S. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. BioMed Research International, 2020. https://doi.org/10.1155/2020/9816142
19. Tharwat, A., Mahdi, H., Elhoseny, M., & Hassanien, A. E. (2018). Recognizing human activity in mobile crowdsensing environment using optimized k-NN algorithm. *Expert Systems with Applications*, 107, 32-44. https://doi.org/10.1016/j.eswa.2018.04.017
20. UPalani Teaching Fellow Professor, Ss. (2022). An IoT Enabled Heart Disease Monitoring System Using Grey Wolf Optimization and Deep Belief Network. https://doi.org/10.21203/rs.3.rs-1058279/v1
21. Wadhawan, S., & Maini, R. (2022). EBPSO: Enhanced binary particle swarm optimization for cardiac disease classification with feature selection. *Expert Systems*, 39(8), 1-20. https://doi.org/10.1111/exsy.13002
22. https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset AD: 02/08/2023

Murat Lüy

Murat Lüy continues as associate professor in Kırıkkale University, Faculty of Engineering and Architecture, Electrical and Electronics Engineering.